

Sensitivity analysis to reduce duplicated features in ANN training for district heat demand prediction

Si Chen^a, Yaxing Ren^a, Daniel Friedrich^b, Zhibin Yu^{a,*}, James Yu^c

^a James Watt School of Engineering, The University of Glasgow, Glasgow G12 8QQ, UK

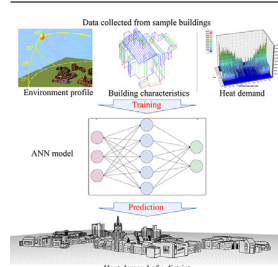
^b School of Engineering, The University of Edinburgh, Edinburgh EH9 3FB, UK

^c SP Distribution PLC, Glasgow G72 0HT, UK

HIGHLIGHTS

- ANN models the relationship between data of weather, building and heat demand.
- Weather-related input variables contain duplicated features. Proposed approach helps ANN model to reduce the training time.
- The approach is validated in a case study of predicting the district heat demand.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 30 July 2020

Received in revised form 10 September 2020

Accepted 11 September 2020

Available online 25 September 2020

Keywords:

Building heat demand prediction

Statistical modelling

Artificial neural network

Sensitivity analysis

Feature selection

ABSTRACT

Artificial neural network (ANN) has become an important method to model the nonlinear relationships between weather conditions, building characteristics and its heat demand. Due to the large amount of training data required for ANN training, data reduction and feature selection are important to simplify the training. However, in building heat demand prediction, many weather-related input variables contain duplicated features. This paper develops a sensitivity analysis approach to analyse the correlation between input variables and to detect the variables that have high importance but contain duplicated features. The proposed approach is validated in a case study that predicts the heat demand of a district heating network containing tens of buildings at a university campus. The results show that the proposed approach detected and removed several unnecessary input variables and helped the ANN model to reduce approximately 20% training time compared with the traditional methods while maintaining the prediction accuracy. It indicates that the approach can be applied for analysing large number of input variables to help improving the training efficiency of ANN in district heat demand prediction and other applications.

1. Introduction

In recent decades, a lot of efforts have made to reduce greenhouse gas emissions and the consumption of fossil fuel to mitigate the global environmental degradation and warming [1]. Many countries have deployment targets for low carbon technologies such as photovoltaic panels, electric vehicles and heat pumps with the aim to cut CO₂ emissions [2]. In terms of global energy consumption, buildings are becoming an important sector in current and future energy landscapes [3–5]. In Europe, the energy consumption of buildings has steadily increased

and accounts for nearly 40% of total energy usage [6]. The data shows that heating, ventilation and air-conditioning (HVAC) systems consume the most electrical energy to provide comfort to the working spaces of the building [7]. In most cold areas, such as Scotland, heating accounts for the highest proportion of total building energy consumption and is much higher than other sectors [8] and nature gas is another significant source of heating. Some local communities, such as universities, have their own combined heat and power (CHP) plants and boilers in the district heat supply for dozens of buildings. The building heat demand can be supplied by either the district heat supply, nature gas, or electricity. Load forecasting is essential for energy consumption management and anomaly detection of energy usage in buildings as well as for the integration of variable renewable sources [9]. But in present, forecasting the

* Corresponding author.

E-mail address: zhibin.yu@glasgow.ac.uk (Z. Yu).

building electricity power consumption is difficult due to the complexity of the system inside the building and varieties of loads, especially the electric appliances [7]. Therefore, the accurate prediction of buildings heat demand, regardless of the supplies of heat demand, becomes the target [10].

The modelling of building heat demand can be classified into three main categories: engineering modelling methods, data-driven methods, and their hybrids [11,12]. Engineering methods develop bottom-up building models and simulate the heat transfer process based on physical principles [13]. Most building energy simulators use bottom-up building models and heat transfer principles to simulate the energy consumption, e.g. TRNSYS, EnergyPlus and Integrated Environmental Solutions Virtual Environment (IES-VE) [14,15]. However, building energy simulators require a detailed description of the building to account for the end-use heat demand and have long simulation times, especially for large energy networks. The required number and accuracy of parameters and computational cost are the main drawbacks of these methods. For these reasons, data-driven methods have become popular as an alternative modelling approach to predict the heat demand of buildings.

The data-driven models are developed using statistical methods to fit the input parameters to outputs without any knowledge of their physical relationship, a so called 'black-box' model. The input parameters for building energy consumption include both the environmental parameters, such as temperature, solar radiation, humidity, and atmospheric pressure, as well as building design parameters, such as percentage area of windows, thermal properties of walls, and building orientation and also the occupant behaviour [16]. Apart from these, recent research indicates that the time series data, such as time of day and days of the week, is also important as input variable of data-driven models [4]. That is because the heating time is normally correlated to both the indoor temperature and heating mode, which can be known as a kind of occupant behaviour. Comparing with the conventional regression modelling technique, artificial intelligent (AI) techniques are known to perform more reliable and efficient in many modelling tasks [17–19]. The commonly used AI techniques include genetic algorithms, support vector machine and artificial neural networks (ANN) [20,21]. ANN has become one of the most important methods in empirical nonlinear modelling and is widely used to model complex functional relationships between weather conditions and building characteristics as inputs and its heat demand as outputs [16,22]. The commonly used measurement methods for validating the result of ANN models include mean absolute error (MAE), mean square error (MSE), correlation (R), and coefficient of determination (R^2) [23]. The major advantages of ANN are its very low model construction cost and ability of flexible input–output mapping for complex systems [24]. Whatever complexity of the target system, the ANN is able to use the simplest construction to model its behaviour. This feature of ANN make it widely used to forecast irregular variables, which includes geography information such as wind speed [25] and global solar radiation [26], random building energy usage such as electricity usage [9], cooling load [27] and energy consumption [17,22], and power generation systems such as photovoltaic [28,29] and PV/thermal system [23].

Although the construction of ANN has these advantages, its shortcomings are the large amounts of data required for the training, long training time, high risk of overfitting and difficulty of interpreting the knowledge gained by 'black-box' models [30]. A common way to reduce these shortcomings is to delete unimportant data components in the training sets to obtain smaller networks, reduced-size data vectors and minimised redundancy in the training data [31,32]. This can be achieved by analysing the total disturbance of network outputs due to perturbed inputs [32]. Reducing the number of inputs to an ANN model and to select key variables is known as feature selection, which aims at identifying the most relevant input features within a dataset [33–35]. Researchers have analysed different methods of feature selection for ranking and identifying important inputs, such as sensitivity analysis, fuzzy curves, and change of MSE [31].

Sensitivity analysis identifies which input parameters are important for the prediction of the output variable and also quantifies how the changes in the values of the input variables alter the value of the output variable [30,36]. Several methods have been proposed to explain the contribution of variables in ANN models, including the adaptation of their connection weights [37], a fictitious input matrix considering a successive variation of one input variable while the others are kept constant [38], the connection weights selected by a randomization approach [39], a perturbation of the input variables [40], the partial derivatives of the output according to the input variables using the connection weights of the ANNs [41,42]. Typical model sensitivity analyses are 'one-at-a-time' simulations that evaluate the impact of each input in turn and ignores the interactions with other input variables [42,43]. However, in predicting the district heat demand with given weather information, it is found that most input variables contain duplicated feature even if they show a high importance in the sensitivity analysis indicating that the training data can be further simplified to reduce the training cost and the risk of overfitting. Therefore, the correlations among inputs also need to be analysed to simplify the training data and remove the duplicated features.

This paper develops a sensitivity analysis method to rank the input variables and to identify input variables with duplicated feature. Both methods are used to remove features in order to reduce the training data and time, and thus improve the efficiency of ANN while maintaining the prediction accuracy. The proposed approach analyses the correlation among inputs by calculating the coefficient of determination of each variable with all others. The results are used to remove variables with low importance as well as variables that have high importance but with duplicated feature from the set of input variables. The approach is evaluated in a case study of predicting the heat demand of a district heating network containing tens of buildings at the campus of the University of Glasgow.

2. Artificial neural network used in building heat prediction

In predicting the energy consumption of a district heating network, using engineering simulation to build a bottom-up model of each building is not efficient because it would require a very large number of hard to get building and occupant activity data, and the simulation would be computationally expensive and time consuming. The motivation of this paper is to build an ANN model that is able to predict different types of buildings. If the model is only trained with the weather profile without giving the building characteristics, it will only be able to predict the heat demand of a particular building. Therefore, the building characteristics is added as the inputs additional to the weather conditions in the ANN training. After the model is well trained, it will be used to predict the heat demand of other buildings by changing the inputs to characteristics of new buildings. Thus, the ANN model is expected to predict the heat demand of a district containing dozens of buildings after trained by the data of environment profile as well as building characteristics and heat demand of some sample buildings, as shown in Fig. 1.

2.1. Artificial neural network model

Neural network is a computational model for nonlinear data fitting that typically includes the input layer, hidden layers, and the output layer [20]. There can be one or more hidden layers depending on the complexity of the model and training data. Each layer has several neurons and every neuron is connected to the output of all the neurons in the previous layer through adaptable synaptic weights [44]. The training of ANN uses a group of input patterns in a data mapping process to produce the dependent variables for the corresponding inputs [45]. In the ANN data mapping process, the neurons in the input layer are multiplied by the weight of corresponding neurons in the hidden layer and then summed up with bias to the neurons in the output layers [22]. The

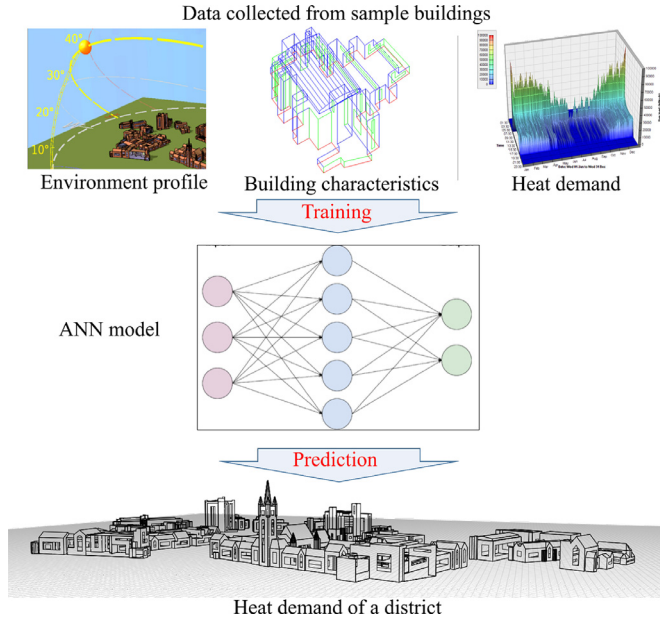


Fig. 1. Structure of neural network to predict the heat demand of a district.

predicted results are compared with the historical data, and their errors are used to update the neurons weights by suitable adaptation [20].

The process of ANN can be described in mathematical formulas. Define $x_k (k = 1, 2, \dots, n)$ as the k th input attribute value which is passed along the links to the other layers. The weighted sum of signals, Σ , arriving at the input of the next neuron is subjected to a transfer function, which is most commonly the *sigmoid* function [46] with the following formula

$$f(\Sigma) = \frac{1}{1 + e^{-\Sigma}} \quad (1)$$

The j th hidden neuron $h_j (j = 1, 2, \dots, p)$ receives the sum of neuron value multiplied by the weights $w_{kj}^{(2)}$ and bias $b_{kj}^{(2)}$ associated with the link as

$$h_j = f\left(\sum_{k=1}^n w_{kj}^{(2)} x_k + b_{kj}^{(2)}\right) \quad (2)$$

The output neurons are defined as $y_i (i = 1, 2, \dots, m)$, which are summed up with their input signals and activation transfer function as

$$y_i = f\left(\sum_{j=1}^p w_{ji}^{(1)} f\left(\sum_{k=1}^n w_{kj}^{(2)} x_k + b_{kj}^{(2)}\right) + b_{ji}^{(1)}\right) \quad (3)$$

where f is the activation function, the sigmoid function used in the paper; $w_{ji}^{(1)}$, $b_{ji}^{(1)}$, $w_{kj}^{(2)}$ and $b_{kj}^{(2)}$ are the weights and bias linked to the output layer (1) and hidden layer (2), respectively. This is a typical two-layer ANN model with an output layer and one hidden layer.

The error between target vector and predicted outputs from ANN model are used to validate the training performance. The appropriate error function is the mean square error (MSE) defined using the differences between the output vector y_i and the target vector t_i as

$$MSE = \frac{1}{m} \sum_{i=1}^m (t_i - y_i)^2 \quad (4)$$

The training error is used to update the ANN parameters of weights and bias of each neurons in the hidden and output layers. The training approaches of ANN include general regression, backpropagation (BP), radial basis function and fuzzy inference system. In this paper, the BP learning algorithm is adopted to a typical two-layer ANN model. The BP algorithm is a supervised iterative training method based on searching the global minimum in the difference between ANN output and target [47]. The errors in the output are propagated back by calculating the

Table 1

Variables and units of input variables for ANN training.

Types	Variables	Units
Times series	Time of a day	h
	Month of a year	month
Environment variables	Dry-bulb temperature	°C
	Wet-bulb temperature	°C
	Dew-point temperature	°C
	Daily mean temperature	°C
	Max. adaptive temp.	°C
	Wind speed	m/s
	Wind direction	deg
	Direct radiation	W/m ²
	Diffuse radiation	W/m ²
	Global radiation	W/m ²
	Solar altitude	deg
	Solar azimuth	deg
	Cloud cover	oktas
	Atmospheric pressure	Pa
	External relative humidity	%
	External moisture content	kg/kg
Building characteristics	Floor area	m ²
	Volume of plant	m ³
	Windows area	%
	Plant radiant fraction	0.0–0.1
	Time of heating	h/day
	Room heating setpoint	°C
	Number of people	pers
	DHW consumption	l/(h · pers)

derivatives that indicate the amount of responsibility of each neuron using the gradient descent method [46]

$$\delta_i^{(1)} = y_i (1 - y_i) (t_i - y_i) \quad (5)$$

$$\delta_j^{(2)} = h_j (1 - h_j) \sum_i \delta_i^{(1)} w_{ji} \quad (6)$$

where $\delta_i^{(1)}$ and $\delta_j^{(2)}$ indicate the responsibilities of output-layer neurons and hidden-layer neurons, respectively. Then the weights and bias of links can be updated based on the responsibilities [48,49] as

$$w_{ji}^{(1)} = w_{ji}^{(1)} + \eta \delta_i^{(1)} h_j \quad (7)$$

$$w_{kj}^{(2)} = w_{kj}^{(2)} + \eta \delta_j^{(2)} x_k \quad (8)$$

$$b_{ji}^{(1)} = b_{ji}^{(1)} + \eta \delta_i^{(1)} \quad (9)$$

$$b_{kj}^{(2)} = b_{kj}^{(2)} + \eta \delta_j^{(2)} \quad (10)$$

where η is the learning rate of the BP neural network.

Apart from the methods and algorithms of machine learning in data-driven models, the data selection with different features and sizes used for training is also a vital factor of the model performance [20].

2.2. Data collection from building energy simulator

The development of a data-driven model will normally consist of data collection and processing as well as model training and testing. The training process of ANN requires a group of datasets from historical data records, which is used as benchmarks to train and test the models performance. The number of inputs neurons depends on the number of input variables for training the ANN. In the original data, the input data has 26 different variables include both the weather conditions and building characteristics as shown in Table 1. The number of hidden neurons is chosen based on the complexity of the target model. The more hidden neurons used, the more complex the model can be achieved, but the longer training time will take. In this paper, the number of hidden neurons is manually chosen considering both the model complexity

and training time. The ANN model is structured with 26 input neurons for ANN training corresponding to the input variables and 20 hidden neurons for internal relation of the ANN model. The data of input variables used for training includes the corresponding hour of the day and months of the year, selected environmental variables, building characteristics and heat demand of all target buildings. The data of building heat demand is used to calculate the training error of the ANN model output.

Ideally, the most direct way is to measure the end-user heat demand of each building. However, there are many difficulties in measuring the real heat demand of the buildings. Firstly, it is required to install great number of sensors in every parts of the buildings. However, as the case study in University of Glasgow (UoG), many buildings are hundreds of years old. These buildings do not have building management system (BMS) as modern buildings and it is difficult to reconstruct them. In addition, the collection of the required historical data will take at least several months but ideally several years in recording data from sensors.

Another way is to measure the heat demand from the supplier. In UoG, the buildings are heated via district heating network supplied by the energy centre. But the energy centre can only record the real heat demand of the whole campus but not each building. It makes difficulties in calculating the heat demand of each building from the sum of heat demand of the whole campus. Therefore, the bottom-up building models are built in the IES-VE software referring to all real buildings in campus. The building parameters are then calibrated to allow the sum of heat demand of all building models matching the real heat demand data of whole campus recorded by the energy centre. IES-VE is an integrated system to build bottom-up models of buildings for thermal analysis and heating load simulation using the Apache engine. After setting the latitude and longitude of the target buildings, the weather profile for the building energy simulation is obtained from the weather station at the nearest airport. For this study it is Glasgow airport which is 9 km from the University of Glasgow. The settings are to make sure the simulated result is as close as possible to the measured heat demand. After the model is well calibrated by the data of whole campus, the simulated heat demand of each building can be obtained. The calibration work and validation result against measured data have been published in previous publications [50,51]. Then the heat demand of each building simulated from this model is trusted to be close to their real heat demand. In this paper, to simplify the data collection process, the training data is taken from simulated results generated in the building energy simulation software IES-VE. The data includes the weather profile, building characteristics and simulated heat demand is used for ANN training.

In addition to the weather profile, different building characteristics will also affect the thermal behaviour of the building. The floor area, building volume, windows area and type, and wall thickness and material are the key characteristics of a building. Apart from these, the building heat demand also depends on its type of operational function. For example, the normal working hours of an office building are from 8:00 to 18:00, and its heat demand is different from that of a restaurant opening from 12:00 to 22:00 or a library open for 24 h. The data is collected from the sample building models built in IES-VE which were calibrated to fit the recorded data of real heat energy consumption [50]. The collected data is normalised before it is used to train the ANN model for the prediction of heat demand of the district heating network.

2.3. Sensitivity analysis of input variables to heat demand

The collected data of weather information and building characteristics are defined as input variables to the ANN model. To validate the effect of model variables, the sensitivity analysis has been designed to provide the elementary effect of each variable. The increase or decrease of each variable will then provide a clear effect on the output. The input and output data used in sensitivity analysis should be exactly the same with that used for ANN training. In the original data, the inputs of sensitivity analysis are the 26 variables including both the environmental

condition and building characteristics, while the outputs are the heat demand of sample buildings randomly chosen from the campus.

The commonly used global sensitivity analysis method is the Morris method, which has been used to test the elementary effect in the case. The basic idea of the Morris method is to evaluate the response of the model output on the basis of a small change in a single input variable. The mean elementary effect of a single variable from the complete set of data points of size U is presented as

$$\mu_k^* = \frac{1}{U} \sum \left| \frac{f(x_1, \dots, x_k + \Delta_k, \dots, x_n) - f(x_1, \dots, x_n)}{\Delta_k} \right| \quad (11)$$

However, in real engineering applications, the input variables are normally varied simultaneously with time and it is difficult to find the effect of a small change of only a single variable on the target variable. Some research in analysing ecology data has realised the potential weakness of perturbing a single factor at a time [30]. The input x_k with n input variables and output $y = f(x_1, \dots, x_n)$ at time step (q) comparing with that at time step ($q-1$) can be presented as

$$x_k^{(q)} - x_k^{(q-1)} = \Delta_k^{(q)} \quad (12)$$

$$y^{(q)} - y^{(q-1)} = f(x_k^{(q-1)} + \Delta_k^{(q)}) - f(x_k^{(q-1)}) \quad (13)$$

Combined with Eq. (11) with $U = 1$, the output of Eq. (13) can be represented by the elementary effect μ_k as

$$\begin{aligned} y^{(q)} - y^{(q-1)} &= \Delta_1^{(q)} \mu_1 + \Delta_2^{(q)} \mu_2 + \dots + \Delta_n^{(q)} \mu_n \\ &= \begin{bmatrix} \Delta_1^{(q)} & \dots & \Delta_n^{(q)} \end{bmatrix} \times \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \end{aligned} \quad (14)$$

In the standard Morris method, the mean value is calculated to average the elementary effect of a single variable in fitting the data. To calculate the elementary effect of multiple variables, the problem is changed to finding a pair of μ_k to fit the relationship between changes on each variable to outputs as

$$\begin{bmatrix} \Delta_1^{(2)} & \dots & \Delta_n^{(2)} \\ \vdots & \Delta_k^{(q)} & \vdots \\ \Delta_1^{(U)} & \dots & \Delta_n^{(U)} \end{bmatrix} \times \begin{bmatrix} \mu_1^* \\ \vdots \\ \mu_n^* \end{bmatrix} = \begin{bmatrix} y(2) - y(1) \\ \vdots \\ y(U) - y(U-1) \end{bmatrix} \quad (15)$$

Normally, the simplest way is to calculate the inverse matrix of $\Delta_k^{(q)}$ and its product with the output. However, the inverse matrix can only be calculated if the original matrix is square and non-singular. This is difficult and not always satisfied in data regression problems. The common approach for the linear equation is to find the least-square solution to minimise the unfitted error referring to all used data points for sensitivity analysis. Then the most suitable μ_k^* can be worked out to solve (15).

2.4. Determine the correlation of input variables

To find the correlation between input variables and outputs, the most common method is regression. Linear regression is a powerful tool that uses mathematical manipulations to transform the relationship between dependent and independent variables into a linear form. Based on this, many procedures were developed to derive the equation of a straight line using the least-squares criterion for calibration. However, most engineering data is poorly represented by a straight line.

An alternative calibration is to fit polynomials to the data using polynomial regression, where the simplest is quadratic regression [52]. The quadratic regression ensures that the first-order derivative is continuous. The least-squares procedure can easily be extended to fit the data to a 2nd-order polynomial as the quadratic least square regression (QLSR) approach. The QLSR has the advantage to integrate both the convergence property of least squares and the probabilistic property of fuzzy regression to fit a non-linear mapping [53–55].

Define the quadratic polynomial equation presented by the input variables as

$$z = a_0 + a_1 s_i + a_2 s_i^2 \quad (16)$$

Assume the fitted value of the k th inputs as x_k can be presented as the sum of polynomial equations of the other $n - 1$ input variables

$$x_k = \sum_{c=1}^{n-1} z_c + \varepsilon \quad (17)$$

where ε is the unique information that is independent with any other input variables. Then, the error between the real and fitted value of each input variable can be represented by

$$e_q = x_k^{(q)} - z_c^{(q)} \quad (18)$$

To calibrate the value of each parameter in the polynomial equation, the estimated value can be obtained by fitting one variable at a time. The effect of other variables can be added to the error. The accumulated square error of U data points can be presented as

$$E_k = \sum_{q=1}^U e_q^2 \quad (19)$$

To minimise the accumulated square error by adjusting the parameters, the partial derivative of the accumulated square error with respect to each parameter can be presented as

$$\begin{cases} \frac{\partial E_k}{\partial a_0} = -2 \sum_{q=1}^U (x_k^{(q)} - a_0 - a_1 s_k^{(q)} - a_2 s_k^{(q)2}) \\ \frac{\partial E_k}{\partial a_1} = -2 \sum_{q=1}^U [(x_k^{(q)} - a_0 - a_1 s_k^{(q)} - a_2 s_k^{(q)2}) s_k^{(q)}] \\ \frac{\partial E_k}{\partial a_2} = -2 \sum_{q=1}^U [(x_k^{(q)} - a_0 - a_1 s_k^{(q)} - a_2 s_k^{(q)2}) s_k^{(q)2}] \end{cases} \quad (20)$$

After defining the partial derivative as zero, the most suitable parameters a_0^* , a_1^* , and a_2^* can be obtained for the quadratic least square regression. After that, the coefficient of determination, R^2 shows the quality of the fit of each variable to the target input as

$$R^2 \equiv 1 - \frac{\sum (x_k - z_c)^2}{\sum (x_k - \bar{x}_k)^2} \quad (21)$$

where

$$\bar{x}_k = \frac{1}{U} \sum_{q=1}^U x_k^{(q)} \quad (22)$$

Based on the coefficient of determination, the variable with the highest R^2 value can be chosen to fit the target variable. Thus, one key variable with its optimised parameters can be chosen in each iteration to fit the target input variable. The remaining difference between the target input variable and its fitted value is used in the next iteration. The fitting process using the QLSR is repeated until the R^2 value is lower than the threshold. The low R^2 value indicates that there has not been enough evidence to show that the remainder is determined by other variables. Then the assumed situation in (17) has been achieved and the process can be stopped. The lower threshold will cause more iterations in fitting the target variable. However, as the error reduced, the excessive iteration times does not give obvious difference in finding the most related features. Thus, considering both the performance and computational load, the threshold is set at 50% in the following case study to have acceptable fitting result. It validates that the correlation among input variables has been found as the target feature can be presented by other input variables. The result can then be used to remove the duplicated features from input variables to reduce the training data.

3. Simulation case study

3.1. Input variables and sensitivity analysis

The data used for ANN training is collected from IES-VE software database containing building characteristics, energy consumption including heat demand, and weather profile, which is recorded by the

nearest weather station in an airport after setting the location. In total, the weather profiles include 16 different hourly recorded variables, as given in Table 1. A number of input variables are shown in Fig. 2 for a period of 30 days.

Fig. 2 (a) shows the five types of temperature information: dry-bulb, wet-bulb, dew-point, daily mean and maximum adaptive temperature. The figures shows that the five temperature variables have similar tendency, i.e. the different temperatures depend on the same weather information. Similar results can also be found in direct radiation, global radiation and diffuse radiation in Fig. 2 (b). They have similar tendency under most circumstance. Fig. 2 (c) shows other useful input weather variables including solar altitude, relative humidity and external moisture content for every hour. They have no obvious dependency on other variables but possibly have hidden and non-linear relationships, which will be discussed later. The ANN training aims to use these weather-based variables along with building characteristics as inputs and corresponding hourly heat demand as outputs.

Due to the multiplicity of the input variables, ANN training could take an extremely long time, furthermore, it reduces the efficiency in finding an effective input-output relationship in the ANN model. The sensitivity analysis (SA) introduced in Section 2.4 aims at analysing the sensitivity of each input variable to the output. The result of SA of input variables is shown in Fig. 3. From the result, the wet-bulb temperature has the highest sensitivity to the building heat demand. The input variables include both the weather-based variables and the building characteristics. From the SA result, the top 5 highly sensitive variables are all temperature related. And the building characteristics with the highest sensitivity for heat demand prediction is the building volume. The wind direction is the variable with the lowest sensitivity to heat demand. With the result of sensitivity analysis, the number of weather input variables can be reduced via choosing the inputs with the highest sensitivity for training.

3.2. Analysis of correlation among input variables

In the sensitivity analysis, the five input variables with the highest influence on the output heat demand are wet-bulb, dry-bulb, dew-point, daily mean and maximum adaptive temperatures. Even though the five variables represent different types of temperature, they contain key information that affects all temperature-based variables. Thus, if the internal relationship between different input variables can be found, the number of input variables can be further reduced.

As the wet-bulb temperature has the highest influence in the sensitivity analysis to the output heat demand, it is tested and used as an example to show the fitting result. The QLSR method proposed in Section 2.3 is used to find the duplicated features in the target input variable. The fitting result in Fig. 4 show the wet-bulb temperature on the y-axis and the value of the correlated variables on the x-axis. Fig. 4 shows the fitting result in the first iteration. From which, it shows the relationship between the wet-bulb temperature with the other fifteen weather-based variables. To make it clearer in comparison, the coefficient of determination, which is also called R-squared (R^2), is used to determine the dependency between variables, as shown in Fig. 5 (a).

The variable with the highest R^2 value is the dry-bulb temperature which has around 97% determination with wet-bulb temperature. After subtracting the determined part of dry-bulb temperature from the wet-bulb temperature, the remaining part is used as the next target to run the second iteration in QLSR. As shown in Fig. 5 (b)–(d), the next determination variables are the relative humidity and dew-point temperature. In the 4th iteration, the R^2 value has dropped below 35% and the error is less than 0.1% of the nominal temperature range. Then the iteration stops.

The fitting result from QLSR shows that the wet-bulb temperature depends on three other features, the dry-bulb temperature, relative humidity and dew-point temperature, whose parameters of weights also provided from the QLSR fitting. After that, the wet-bulb temperature

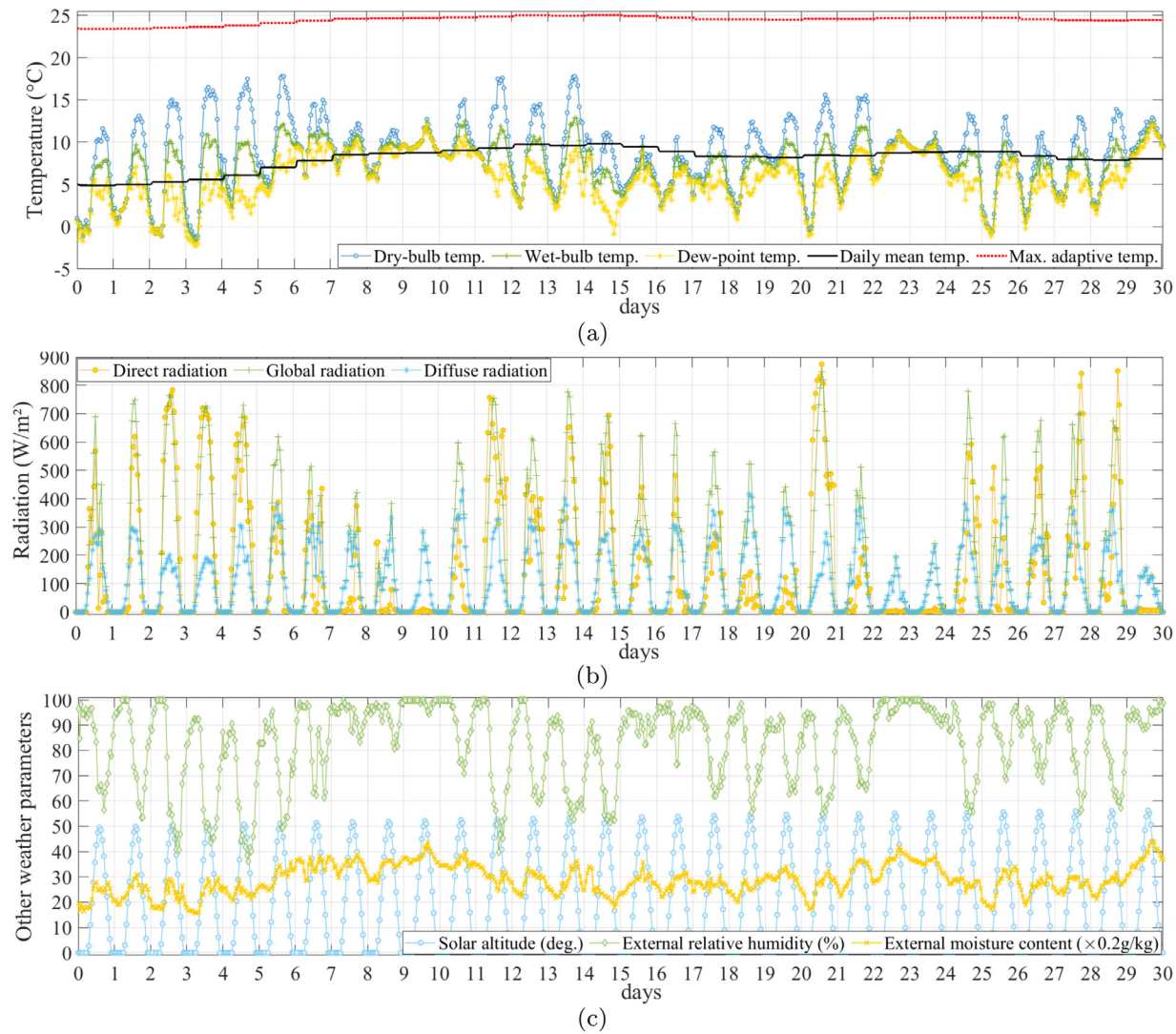


Fig. 2. A sample of input variables used in ANN training: (a) dry-bulb, wet-bulb, external dew-point, daily mean and max adaptive temperatures, (b) direct, global and diffuse radiations, (c) solar altitude, external relative humidity, and moisture content

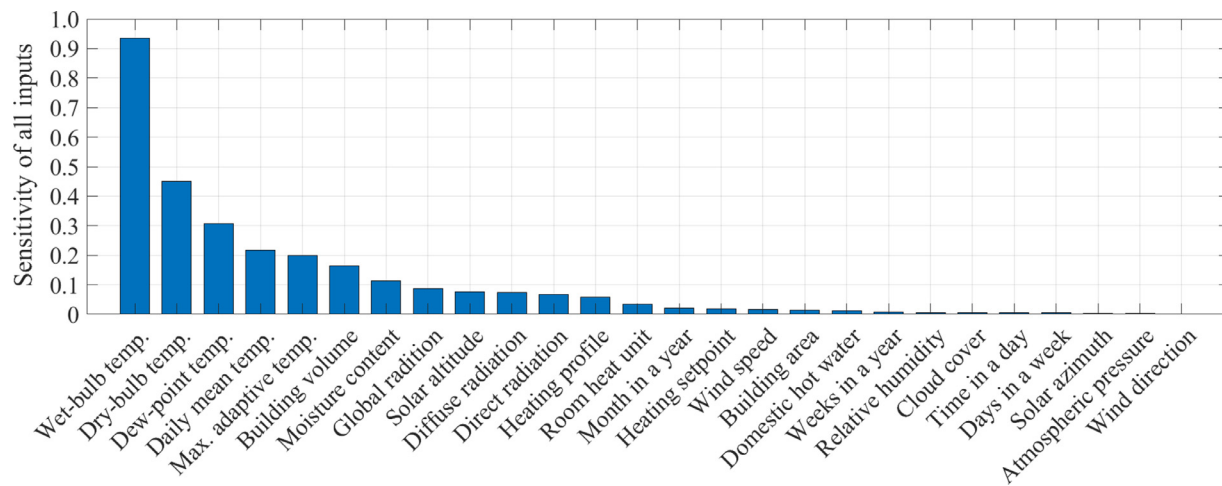


Fig. 3. Sensitivity analysis result of all input variables

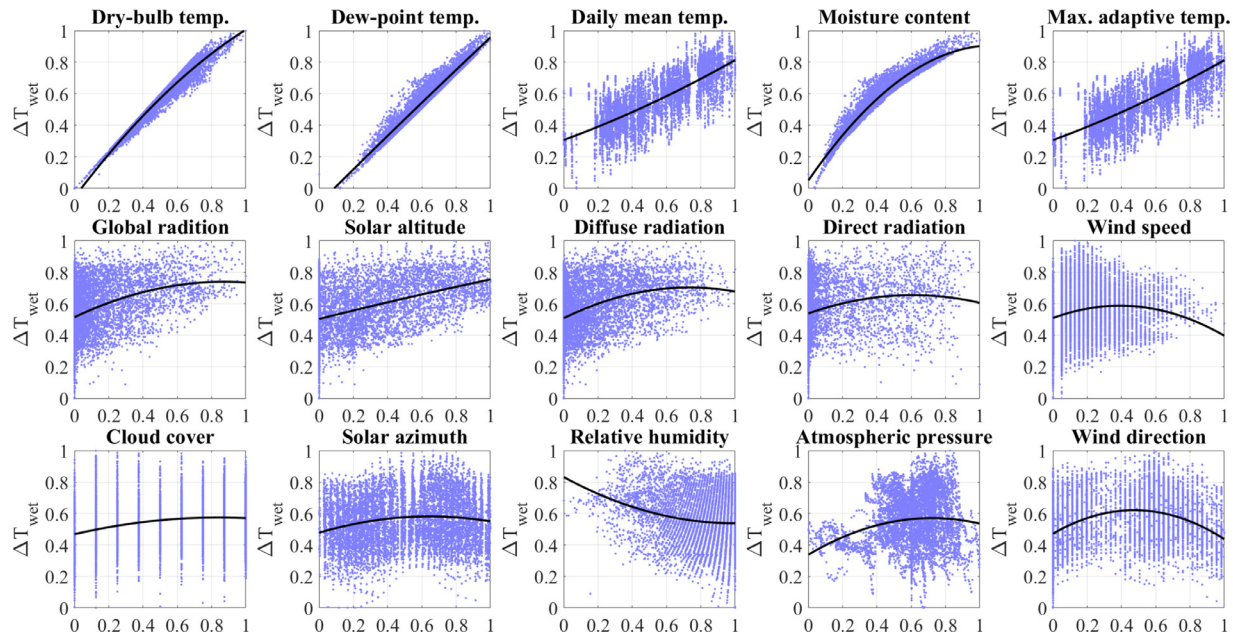


Fig. 4. Quadratic least square regression fitting curve of the wet-bulb temperature using other weather-based input variables (1st iteration).

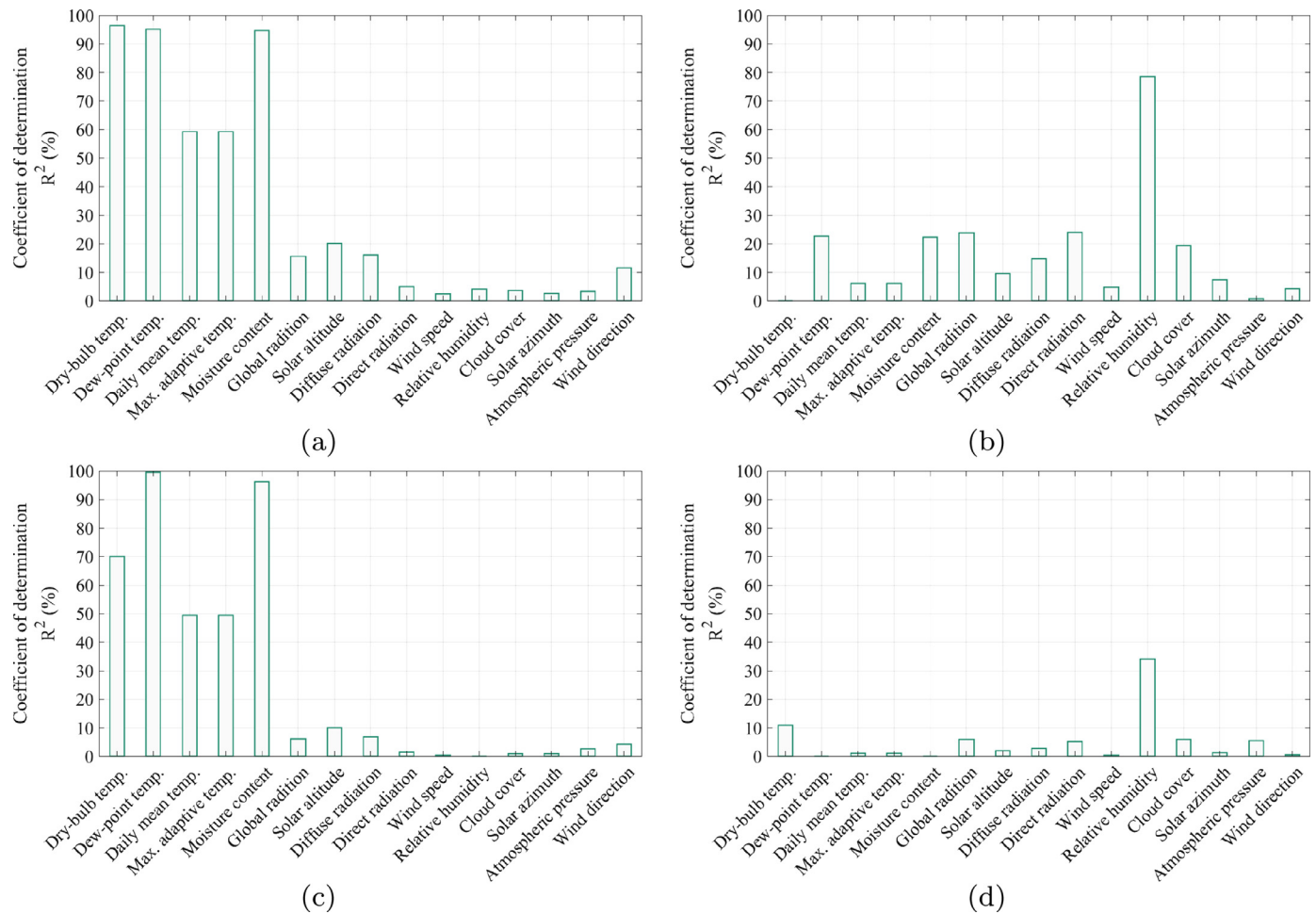


Fig. 5. Coefficient of determination of other weather-based input variables to Wet-bulb temperature in each iteration: (a) 1st iteration, (b) 2nd iteration, (c) 3rd iteration and (d) 4th iteration.

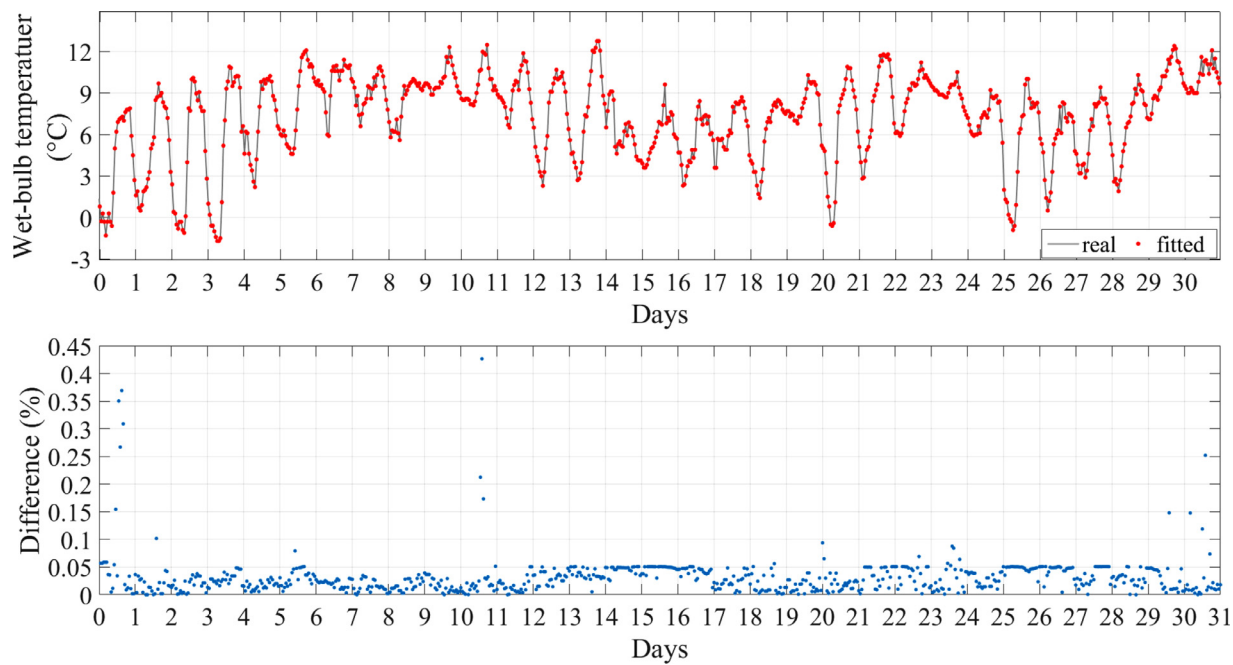


Fig. 6. Fitting result of real wet-bulb temperature using searched input variables: (a) comparison between real and fitted wet-bulb temperature, (b) difference between real and fitted temperature.

can be fitted by other three and the fitting weights based on equation (16). The fitted wet-bulb temperature is compared with its real value as demonstrated in Fig. 6. In the figure, the solid black line indicates the real wet-bulb temperature and the red dots indicate the fitted one that is calculated by other features. The result shows that the average fitting error, which is given in the bottom figure with the blue dots, is less than 0.5%. It also means that the three fitted variables have more than 99.5% information of wet-bulb temperature. Therefore, it verifies that the wet-bulb temperature can be fully described by the other three features.

If the wet-bulb temperature and other three features are all chosen as the input variables for the ANN training, the output heat demand will be found related to both the wet-bulb temperature and the other three features. It causes the repeated training to the duplicated features. Thus, in order to improve the effectiveness of ANN training, the duplicated feature of wet-bulb temperature can be removed from the training inputs and the number of variables are then reduced for a faster training speed.

Similar with the QLSR approach used on the wet-bulb temperature fitting, other variables are also tested for duplicated features before they are used as inputs of ANN training. In the traditional sensitivity analysis result given in Fig. 3, the cut-off criteria is set as 0.02 to allow 15 variables out of 26 used for the ANN training. After running the same approach on the 15 variables, the result indicates that 3 features out of the 15 can be further removed to reduce the training load of ANN. They are the wet-bulb temperature, the maximum adaptive temperature and the moisture content. Using the QLSR approach, they are found containing duplicated features with other variables and thus can be removed from the training inputs. All other 12 variables, which are known as the result of sensitivity analysis with reduced features, are chosen for the ANN training.

However, it has to say that the number of variables with duplicated features depends on collected data and different case studies. In other cases that using different data sources, the removed variables could be found no longer containing duplicated feature with others. Thus, it is necessary to run the proposed sensitivity analysis with reduced features approach for each data source and case study.

3.3. ANN training and prediction performance comparison

To verify the effectiveness of the proposed approach, this section gives the comparison among the ANN training result of using all 26 input variables (All), the top 15 input variables from sensitivity analysis (SA), and 12 input variables chosen from the sensitivity analysis with reduced features (SARF).

The ANN is built and trained in Matlab with the built-in neural network toolbox. The parameter of ANN models is set using 1 hidden layer with 20 hidden nodes. The models are from the building models from the University of Glasgow, which has 36 different buildings with different heating types, and the total recorded data of 12 months weather profile. The training of ANN uses part of the recorded data includes weather profile of 4 months from January to April, building characteristics of 10 randomly chosen sample buildings and their heat demand. As the weather profile and heat demand of the first 4 months and 10 sample buildings are used to train the ANN, the remained 8 months weather profile is used to test the training performance by predicting the heat demand of the remained 26 buildings.

Furthermore, the ANN training is using random initial weights and bias value. In the ANN training, the stopping criteria is that the gradient reduction of training error using the initial parameters is low enough, which means the training error is difficult to be further reduced. The training performance is mostly dependent on the initial parameters of weights and bias. Therefore, only one set of training result is not enough to show and compare the performance for randomly chosen initial weights. The simulation test uses the MATLAB neural network toolbox for ANN model training. In order to make a fair comparison and to give a convincing conclusion, the ANN training and heat demand prediction should be repeated many times so that the statistic results can be compared to show the average training performance. In the case study, each method for ANN training has been repeated one thousand times and the statistical results of all methods are compared. The probability density functions of statistical prediction error of heat demand and training time of ANN model for the one thousand repeated tests are shown in Fig. 7.

Due to the randomness of the ANN training process, the results show some variability and, thus, the medians of the statistical results are given

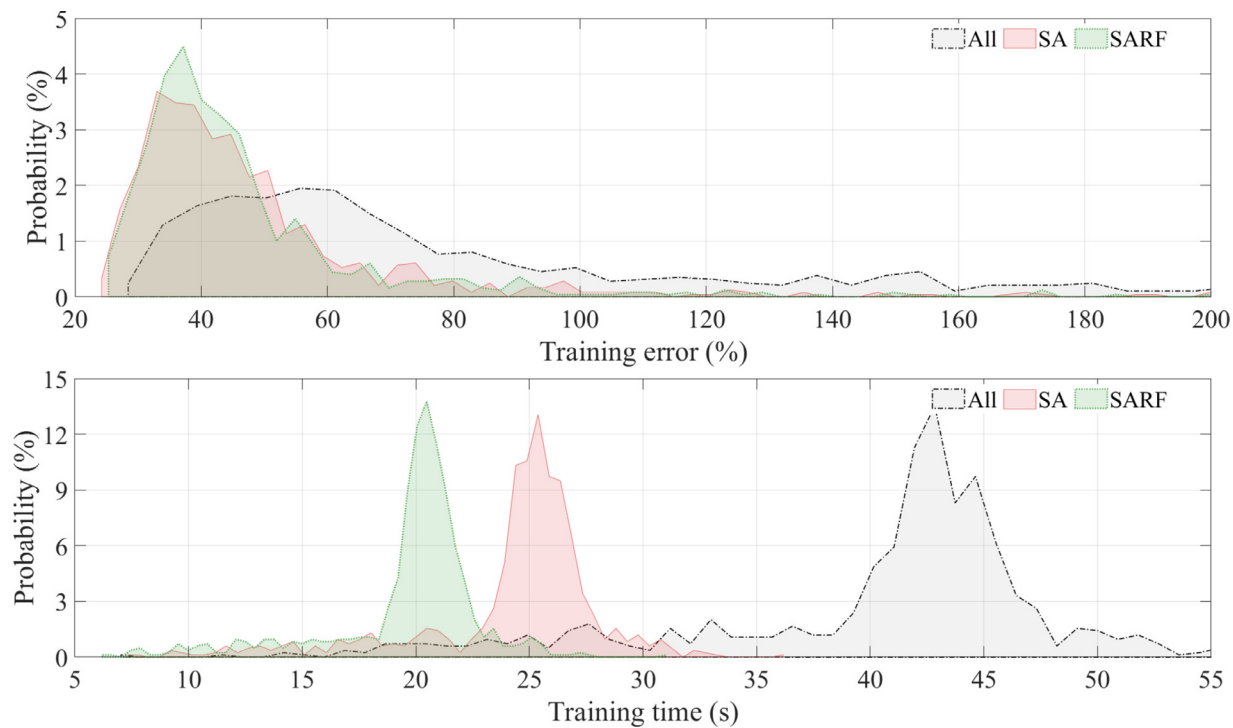


Fig. 7. Statistic ANN training performance and result comparison: (a) probability of training error based on statistics, (b) probability of training time based on statistics.

Table 2

ANN test result in heat demand prediction using original 26 input variables.

Algorithm	Mean error (%)	Min. error (%)	Max. error (%)	Std. (%)	Mean time (s)	Min. time (s)	Max. time (s)	Std. (s)
LM	88.81	24.56	199.96	52.99	42.47	33.46	58.63	5.93
BFG	29.90	20.56	121.61	8.38	16.77	13.47	23.38	1.54
RP	33.66	21.58	112.60	10.94	4.05	3.10	5.81	0.43
SCG	29.41	21.34	93.62	7.81	7.06	5.23	8.90	0.69
CGB	29.52	20.25	83.62	8.07	14.61	7.57	21.86	1.58
CGF	29.53	21.43	112.18	7.29	14.23	10.84	18.27	1.55
CGP	29.59	19.99	86.54	7.06	14.37	0.69	19.29	1.59
OSS	30.39	21.93	89.76	6.31	13.12	9.70	16.36	1.36
GDX	63.18	38.80	171.49	13.58	4.06	3.21	5.77	0.43

Table 3

ANN test result in heat demand prediction using 15 input variables from SA.

Algorithm	Mean error (%)	Min. error (%)	Max. error (%)	Std. (%)	Mean time (s)	Min. time (s)	Max. time (s)	Std. (s)
LM	42.02	20.21	195.74	27.84	24.81	27.17	57.70	5.09
BFG	25.46	18.06	46.96	3.88	15.53	12.05	19.80	1.78
RP	29.09	20.84	58.42	4.89	4.05	2.99	5.74	0.52
SCG	25.72	19.29	55.29	3.64	6.96	5.50	11.67	0.89
CGB	25.29	18.98	54.58	3.84	14.67	0.51	22.86	2.24
CGF	25.40	18.98	43.80	3.42	14.44	10.83	21.45	1.87
CGP	25.66	19.18	49.98	3.96	14.55	11.25	22.73	2.12
OSS	27.74	20.66	55.91	3.88	13.15	10.42	19.23	1.64
GDX	59.32	33.85	153.99	15.28	4.12	3.08	8.18	0.65

and compared. In the result of Fig. 7 (a), the median of prediction error using the original method is about approximate 60% prediction error while that of SA and SARF are around 38% prediction error. And in Fig. 7 (b), it is clear that the median of ANN training time with all variables is about 43 s, the median of training time of SA is 26 s and that of SARF is 21 s, around 20% less training time than SA and 50% less than original all inputs. In addition to the median of prediction error and training time, the variance of SARF is obviously smaller than the variance of the case with all input variables and similar to the variance of SA. This verifies that SARF will reduce the uncertainty in ANN training and training time while achieving better performance.

In addition, the result in Fig. 7 shows that the ANN training performance using the Levenberg–Marquardt (LM) training function, which is the most widely used training algorithm. However, there are many other training algorithms, including BFGS Quasi-Newton (BFG), resilient backpropagation (RP), scaled conjugate gradient (SCG), conjugate gradient with Powell/Beale restarts (CGB), Fletcher–Powell conjugate gradient (CGF), Polak–Ribière conjugate gradient (CGP), one step secant (OSS), and variable learning rate backpropagation (GDX). The next step is to verify the performance of the developed SARF method in other training algorithms. Tables 2–4 show the prediction error of all training functions using the original 26 input variables, 15 input variables

Table 4
ANN test result in heat demand prediction using 12 input variables from SARF.

Algorithm	Mean error (%)	Min. error (%)	Max. error (%)	Std. (%)	Mean time (s)	Min. time (s)	Max. time (s)	Std. (s)
LM	38.31	20.84	209.48	21.53	20.80	3.95	38.34	7.71
BFG	25.72	17.48	38.74	3.08	12.65	9.15	19.87	1.75
RP	29.60	22.08	54.95	4.86	3.35	2.69	4.83	0.44
SCG	26.02	18.72	35.00	3.07	6.29	4.98	9.70	0.84
CGB	25.43	19.03	41.73	3.16	12.24	7.86	17.68	1.66
CGF	26.13	19.06	38.25	3.12	11.97	9.69	17.53	1.48
CGP	26.37	19.87	48.80	3.62	12.02	9.22	18.94	1.84
OSS	28.27	20.96	43.46	3.68	10.95	8.92	17.13	1.45
GDX	60.31	39.64	119.06	14.07	3.34	1.37	5.23	0.53

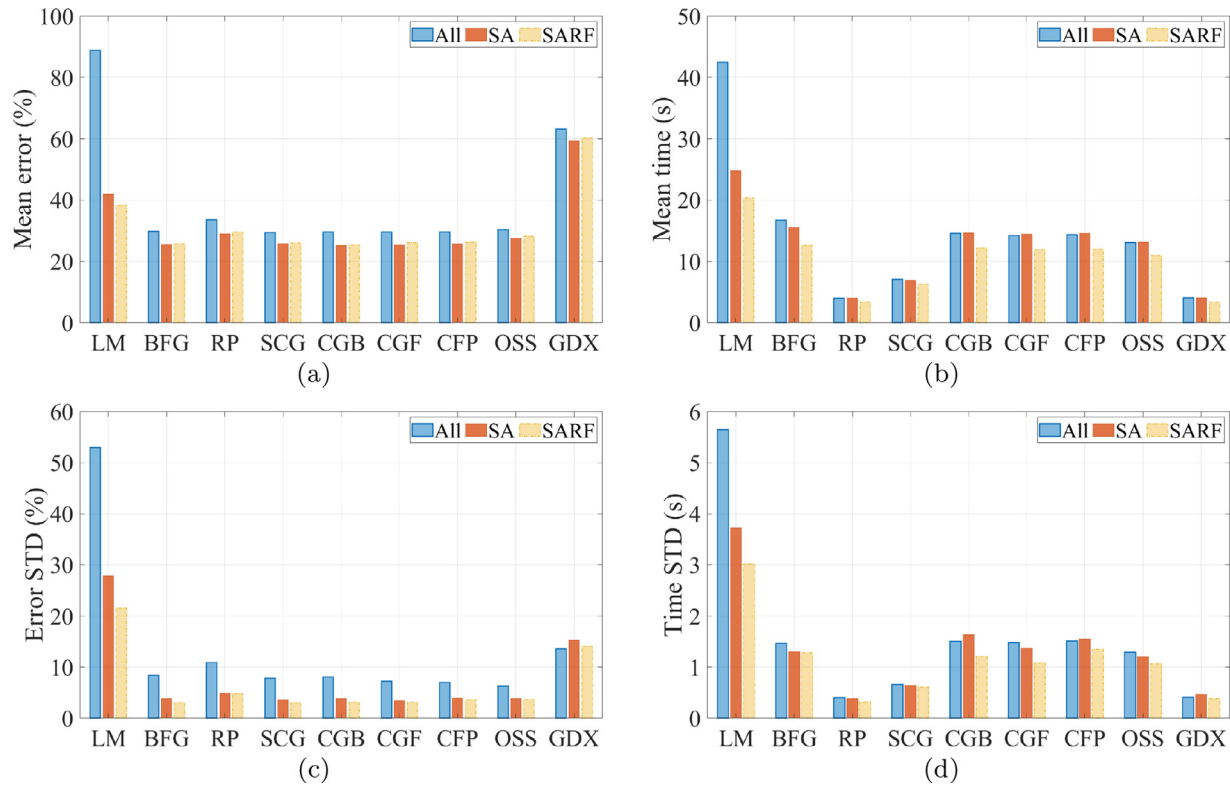


Fig. 8. Comparison of using SARF with traditional SA method in different training functions. (a) Mean prediction error, (b) mean training time, (c) standard deviation of prediction error, (d) standard deviation of training time.

from SA, and 12 input variables from SARF. The indices for performance comparison are chosen as the mean, min and max prediction error its standard deviation (STD) as well as the mean, min and max prediction time and its STD.

To make the comparison clearer, the mean prediction error, error STD, mean training time, and time STD are drawn in bar charts as shown in Fig. 8. The results show the prediction error and time of different training functions in ANN model training. In comparison among different training functions, the BFG, SCG, CGB, CGG, and CGP have the relatively better prediction result while the RP and GDX require the least training time. For all the training function used in the above ANN training, the SARF uses the least training time to perform a relatively better performance with the least prediction error. It verifies that the proposed SARF method can improve the ANN training efficiency and reduce the training time to obtain the same or even better performance in the heat demand prediction.

4. Conclusion

With the development of ANN technology, the sensitivity analysis is necessary to rank the importance of input variables due to a large

amount of training data. In predicting the district heat demand using weather information, it is found that most input variables contain duplicated features which is not required to train the ANN model. This paper proposed a method with the ability to remove both the variables with low importance and the variables that have high importance but contain duplicated features. The proposed approach analysed the correlation among input variables via detecting the coefficient of determination of each variable with others referring to the fitting error of quadratic least square regression. The approach is validated in a case study of predicting heat demand in a district using an ANN model that is trained by historic data from several sample buildings. The traditional sensitivity analysis method ranked the input variables based on their influence on the heat demand. It was shown that the 15 most important features can be used to predict the district heat demand with the same or even better performance than the complete set of features. The proposed method further removed 3 important variables that are determined by other variables via analysing the determination of each variable. The results show that the proposed method can reduce training time by around 20% while achieving the same training and prediction performance compared with the traditional sensitivity analysis method. With the developed sensitivity and correlation analysis approach, the

training data is simplified and the efficiency of training an ANN model can then be improved.

Declaration of Competing Interest

The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

Acknowledgment

The research presented in this article was undertaken as part of a project joint founded by Energy Technology Partnership (ETP), SP Distribution PLC (Scottish Power), grant number 146. The authors would like to thank Gillian Brown for providing valuable information on the campus energy consumption.

References

- [1] Heo Y, Choudhary R, Augenbroe G. Calibration of building energy models for retrofit analysis under uncertainty. *Energy Build* 2012;47:550–60.
- [2] Lyons P, Wade N, Jiang T, Taylor P, Hashiesh F, Michel M, et al. Design and analysis of electrical energy storage demonstration projects on UK distribution networks. *Appl Energy* 2015;137:677–91.
- [3] De Rosa M, Bianco V, Scarpa F, Tagliafico LA. Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. *Appl Energy* 2014;128:217–29.
- [4] Jang J, Baek J, Leigh S-B. Prediction of optimum heating timing based on artificial neural network by utilizing BEMS data. *J Build Eng* 2019;22:66–74.
- [5] Reynolds J, Ahmad MW, Rezgui Y, Hippolyte J-L. Operational supply and demand optimisation of a multi-vector district energy system using artificial neural networks and a genetic algorithm. *Appl Energy* 2019;235:699–713.
- [6] Zhao H-x, Magoulès F. A review on the prediction of building energy consumption. *Renew Sustain Energy Rev* 2012;16(6):3586–92.
- [7] Ahmad A, Hassan M, Abdullah M, Rahman H, Hussin F, Abdullah H, et al. A review on applications of ann and SVM for building electrical energy consumption forecasting. *Renew Sustain Energy Rev* 2014;33:102–9.
- [8] Ozkahraman HT, Bolattürk A. The use of tuff stone cladding in buildings for energy conservation. *Constr Build Mater* 2006;20(7):435–40.
- [9] Chae YT, Horesh R, Hwang Y, Lee YM. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build* 2016;111:184–94.
- [10] Wei Y, Xia L, Pan S, Wu J, Zhang X, Han M, et al. Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. *Appl Energy* 2019;240:276–94.
- [11] Fouquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: a review. *Renew Sustain Energy Rev* 2013;23:272–88.
- [12] Fumo N. A review on the basics of building energy estimation. *Renew Sustain Energy Rev* 2014;31:53–60.
- [13] Balaji N, Mani M, Reddy BV. Dynamic thermal performance of conventional and alternative building wall envelopes. *J Build Eng* 2019;21:373–95.
- [14] Connolly D, Lund H, Mathiesen BV, Leahy M. A review of computer tools for analysing the integration of renewable energy into various energy systems. *Appl Energy* 2010;87(4):1059–82.
- [15] Hay S., Ferguson A. A review of power system modelling platforms and capabilities. IET Special Interest Publication for the Council for Science and Technology 2015;3–13.
- [16] Ekici BB, Aksoy UT. Prediction of building energy consumption by using artificial neural networks. *Adv Eng Softw* 2009;40(5):356–62.
- [17] Martellotta F, Ayr U, Stefanizzi P, Sacchetti A, Riganti G. On the use of artificial neural networks to model household energy consumptions. *Energy Proc* 2017;126:250–7.
- [18] Hunter A, Kennedy L, Henry J, Ferguson I. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Comput Methods Programs Biomed* 2000;62(1):11–19.
- [19] Haykin S. *Neural networks: a comprehensive foundation*. Prentice-Hall, Inc.; 2007.
- [20] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. *Renew Sustain Energy Rev* 2018;81:1192–205.
- [21] Bai H, Liu C, Breaz E, Gao F. Artificial neural network aided real-time simulation of electric traction system. *Energy AI* 2020:100010.
- [22] Biswas MR, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. *Energy* 2016;117:84–92.
- [23] Al-Waeli AH, Sopan K, Yousif JH, Kazem HA, Boland J, Chaichan MT. Artificial neural network modeling and analysis of photovoltaic/thermal system based on the experimental study. *Energy Convers Manag* 2019;186:368–79.
- [24] Pastor-Bárcenas O, Soria-Olivas E, Martín-Guerrero JD, Camps-Valls G, Carrasco-Rodríguez JL, del Valle-Tascón S. Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling. *Ecol Model* 2005;182(2):149–58.
- [25] Noorollahi Y, Jokar MA, Kalhor A. Using artificial neural networks for temporal and spatial wind speed forecasting in Iran. *Energy Convers Manag* 2016;115:17–25.
- [26] Alsina EF, Bortolini M, Gamberi M, Regattieri A. Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Convers Manag* 2016;120:320–9.
- [27] Deb C, Eang LS, Yang J, Santamouris M. Forecasting diurnal cooling energy load for institutional buildings using artificial neural networks. *Energy Build* 2016;121:284–97.
- [28] Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Math Comput Simul* 2017;131:88–100.
- [29] Dolara A, Grimaccia F, Leva S, Mussetta M, Ogliari E. A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies* 2015;8(2):1138–53.
- [30] Gevrey M, Dimopoulos I, Lek S. Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecol Model* 2006;195(1-2):43–50.
- [31] Sung A. Ranking importance of input parameters of neural networks. *Exp Syst Appl* 1998;15(3-4):405–11.
- [32] Zurada JM, Malinowski A, Cloete I. Sensitivity analysis for minimization of input data dimension for feedforward neural network. In: *Proceedings of the IEEE international symposium on circuits and systems-ISCAS'94*, 6. IEEE; 1994. p. 447–50.
- [33] Srivastava L, Singh S, Sharma J. Comparison of feature selection techniques for ann-based voltage estimation. *Electr Power Syst Res* 2000;53(3):187–95.
- [34] Crone SF, Kourentzes N. Feature selection for time series prediction—a combined filter and wrapper approach for neural networks. *Neurocomputing* 2010;73(10-12):1923–36.
- [35] Li Z, Friedrich D, Harrison GP. Demand forecasting for a mixed-use building using agent-schedule information with a data-driven model. *Energies* 2020;13(4):780.
- [36] Blower SM, Dowlatabadi H. Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. *Int Stat Rev/Revue Internationale de Statistique* 1994:229–43.
- [37] Goh AT. Back-propagation neural networks for modeling complex systems. *Artif Intell Eng* 1995;9(3):143–51.
- [38] Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S. Application of neural networks to modelling nonlinear relationships in ecology. *Ecol Model* 1996;90(1):39–52.
- [39] Olden JD, Jackson DA. Illuminating the black box: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Model* 2002;154(1-2):135–50.
- [40] Scardi M. Artificial neural networks as empirical models for estimating phytoplankton production. *Mar Ecol Progr Ser* 1996;139:289–99.
- [41] Dimopoulos Y, Bourret P, Lek S. Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Process Lett* 1995;2(6):1–4.
- [42] Henderson-Sellers B, Henderson-Sellers A. Sensitivity evaluation of environmental models using fractional factorial experimentation. *Ecol Model* 1996;86(2-3):291–5.
- [43] Beres DL, Hawkins DM. Plackett–Burman technique for sensitivity analysis of many-parametered models. *Ecol Model* 2001;141(1-3):171–83.
- [44] Kumar R, Aggarwal R, Sharma J. Energy analysis of a building using artificial neural network: a review. *Energy Build* 2013;65:352–8.
- [45] Magoulès F, Zhao H-X. *Data mining and machine learning in building energy analysis*. Wiley Online Library; 2016.
- [46] Kubat M. *An introduction to machine learning*. Springer; 2017.
- [47] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986;323(6088):533–6.
- [48] Attoh-Okine NO. Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Adv Eng Softw* 1999;30(4):291–302.
- [49] Ye Z, Kim MK. Predicting electricity consumption in a building using an optimized back-propagation and Levenberg–Marquardt back-propagation neural network: case study of a shopping mall in china. *Sustain Cities Soc* 2018;42:176–83.
- [50] Chen S, Friedrich D, Yu Z, Yu J. District heating network demand prediction using a physics-based energy model with a Bayesian approach for parameter calibration. *Energies* 2019;12(18):3408.
- [51] Chen S, Ren Y, Friedrich D, Yu Z, Yu J. District heat demand prediction using artificial neural network with data of sample building. In: *Proceedings of the international conference on applied energy (ICAE 2019)*; 2019. August 12-15, Västerås, Sweden.
- [52] Gu H, Liu G, Wang J, Aubry A-F, Arnold ME. Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance. *Anal Chem* 2014;86(18):8959–66.
- [53] Tanaka H, Lee H. Interval regression analysis by quadratic programming approach. *IEEE Trans Fuzzy Syst*. 1998;6(4):473–81.
- [54] Baffi G, Martin E, Morris A. Non-linear projection to latent structures revisited: the quadratic PLS algorithm. *Comput Chem Eng* 1999;23(3):395–411.
- [55] Baffi G, Martin E, Morris A. Non-linear projection to latent structures revisited (the neural network PLS algorithm). *Comput Chem Eng* 1999;23(9):1293–307.